

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
19 February 2004 (19.02.2004)

PCT

(10) International Publication Number  
**WO 2004/015686 A1**

(51) International Patent Classification<sup>7</sup>: **G10L 15/20**,  
15/18

(21) International Application Number:  
PCT/EP2002/008585

(22) International Filing Date: 1 August 2002 (01.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)**  
[SE/SE]; S-126 25 Stockholm (SE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SCHLEIFER, Ralph** [DE/DE]; Fronhofstrasse 12a, 91257 Pegnitz (DE). **KIESSLING, Andreas** [DE/DE]; Atzelsberger Strasse 12, 91080 Marloffstein (DE). **HIRSCH, Hans-Günter** [DE/DE]; Weissdornweg 12, 52249 Eschweiler (DE).

(74) Agents: **WITTRUP, Flemming** et al.; Zacco Denmark A/S, Hans Bekkevolds Allé 7, DK-2900 Hellerup (DK).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,

CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

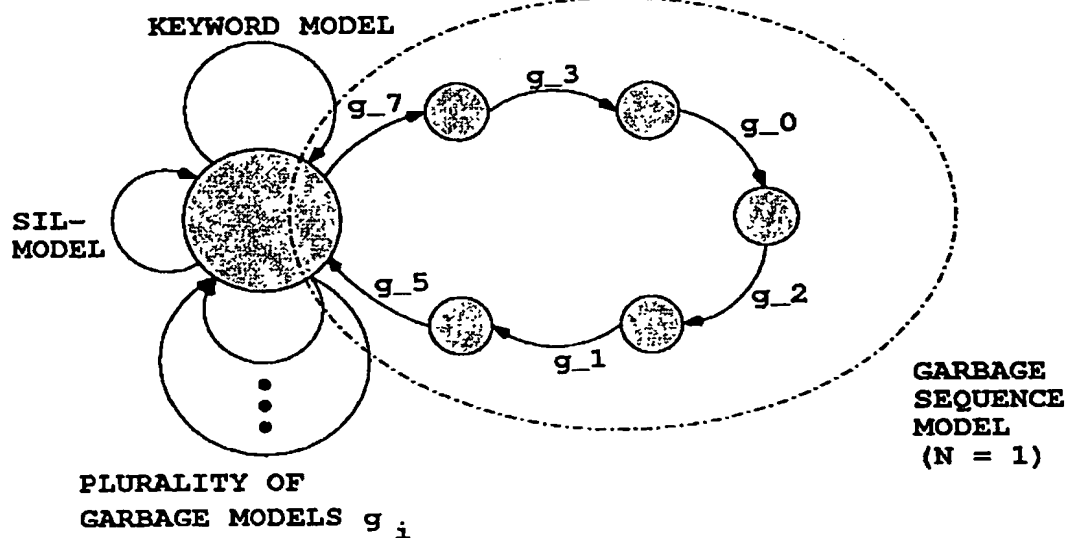
(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent

[Continued on next page]

(54) Title: METHOD FOR AUTOMATIC SPEECH RECOGNITION



(57) Abstract: A method for recognizing a keyword from a spoken utterance is based on at least one keyword model and a plurality of garbage models. Then a part of the spoken utterance is assessed as the keyword to be recognized, if that part matches best either to the keyword model or to a garbage sequence model. Here, the garbage sequence model is a series of consecutive garbage models from that plurality of garbage models.



(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**Published:**

— with international search report

## Method for Automatic Speech Recognition

The present invention relates to a method for automatic speech recognition. In particular the present invention relates to a  
5 method for recognizing a keyword from a spoken utterance.

A method for automatic speech recognition, where a single or a plurality of keywords is recognized in a spoken utterance, is often named as keyword spotting. For each keyword to be  
10 recognized, a keyword model is trained and stored. Each keyword model is trained either for speaker dependent or speaker independent speech recognition and represents for example a word or a phrase. A keyword is spotted from the spoken utterance, when the spoken utterance itself or a part thereof  
15 matches best to any of the previously created and stored keyword models.

In the recent years, such a method for speech recognition often has been used in mobile equipment, like e.g. in mobile phones. With it, the mobile equipment can be partly or fully  
20 controlled with voice commands instead of using the keyboard. The method is preferably useable in car hands-free equipment, where it is forbidden to handle the mobile phone with the keyboard. Hereby, the mobile phone is activated as soon as a keyword is determined from a spoken utterance of the user.  
25 Then, the mobile phone listens for a further spoken utterance and assesses parts thereof as the keyword to be recognized, if that part matches best to any of the stored keyword models.

Depending on the acoustic environment, where the mobile equipment is used, or depending on the users behaviour, like  
30 e.g. the pronunciation, the keywords are recognized more or less correctly. For example, the assessing could be wrong, if the part of the spoken utterance is matched to one of the

stored keywords, but which is not the wanted keyword to be recognized. As a consequence, the hit rate, that is the number of correctly recognized keywords relative to the total number of spoken keywords, strongly depends on the acoustic environment and the users behaviour.

Methods for automatic speech recognition, known from prior art, often use so called garbage models in addition to the keyword models [A new approach towards Keyword Spotting, Jean-Marc Boite, EUROSPEECH Berlin, 1993, pp.1273-1276]. For this, a plurality of garbage models is created. Some garbage models represent for example non-keyword speech, like lip smacks, breaths, or filler words "aeh" or "em". Other garbage models are created to represent background noise. The garbage models are e.g. phonemes, phoneme cover classes, or complete words. By utilising these garbage models, the false alarm rate, that is the number of wrongly recognized keywords per time unit, is decreased. That is, because parts of the spoken utterance, which include non-keyword speech can be mapped directly to one of the stored garbage models. But, when applying such a method, the hit rate is decreased, because a part of the spoken utterance might matches better to one or more of the plurality of garbage models, than to the keyword model itself. For example, if during the recognition phase the acoustic environment is bad, the part of the spoken utterance might matches to a garbage model, which represents such an acoustic environment. As a result, that part is assessed as non-keyword speech, which is of course not the wanted result.

It is therefore the object of the present invention to provide a method for speech recognition, which increases the hit rate and avoids the disadvantages of the known prior art.

This is solved by the method of claim 1. According to the present invention, there is provided a method for recognizing a keyword from a spoken utterance, with at least one keyword model and a plurality of garbage models, wherein a part of the spoken utterance is assessed as a keyword to be recognized, if that part matches best either to the keyword model or to a garbage sequence model, and wherein the garbage sequence model is a series of consecutive garbage models from that plurality of garbage models.

Essentially, then the method of the present invention also assessed a part of a spoken utterance as a keyword to be recognized, when that part of the spoken keyword matches best to the garbage sequence model. Then, as an advantage of the present invention, the hit rate is increased. That is, because two models, the keyword model and the garbage sequence model, are used to recognize the keyword from a spoken utterance. Here, in the context of the present invention, a part of the spoken utterance is any time interval of an incoming utterance. The length of the time interval can be the complete utterance or only a small sequence thereof.

Advantageously, the method in accordance with the present invention avoids that the hit rate is decreased, when garbage models exist, which, in series, match better to the spoken utterance than the keyword model itself. Therefore the present automatic speech recognition method is more robust than known prior art speech recognition methods.

Preferably the garbage sequence model is determined by comparing a keyword utterance, which represents the keyword to be recognized with the plurality of garbage models, and detecting the series of consecutive garbage models, which match best to the keyword. With it, the garbage sequence model is easily created, based on existing garbage models as already used for prior art speech recognition methods. Such a prior art

method is e.g. based on a finite state syntax, where one or more keyword models and a plurality of garbage models are used to recognize keywords from any incoming utterance. According to the present invention, the garbage sequence model is then  
5 created with a finite state syntax, which only includes the plurality of garbage models, but not the keyword models. The incoming utterance, which is the keyword utterance and represents the keyword, is compared with the plurality of stored garbage models. Then a series of consecutive garbage  
10 models from the plurality of garbage models is determined as the garbage sequence model, which best represent the keyword. According to the present invention this garbage sequence model is then used to recognize the keyword from a spoken utterance, if a part of the spoken utterance matches either to the keyword  
15 model or to that determined garbage sequence model.

In accordance with the method of the present invention, the determined garbage sequence model is privileged against any other path through the plurality of garbage models. Especially, the determined garbage sequence model is privileged against any  
20 path, which includes the same series of consecutive garbage models. This provides, that the part of the spoken utterance is assessed as the keyword to be recognized, although a similar path through the plurality of garbage models exists. Therefore, the hit rate is increased, because then the part of the spoken  
25 utterance is preferably assessed as the keyword to be recognized.

In accordance with a first aspect of the present invention, further, a number of further garbage sequence models is determined, which also represent that keyword, and the part  
30 of the spoken utterance is assessed as the keyword to be recognized, if that part of the spoken utterance matches best to any of that number of garbage sequence models. Then a total number of garbage sequence models, and the keyword model are used to recognize the keyword. With it, the hit rate is

increased, because also a slightly worse spoken utterance might matches to any of the further garbage sequence models and is therefore assessed as the keyword.

The total number of garbage sequence models is preferably  
5 determined, by calculating for each garbage sequence model a probability value and selecting those garbage sequence models as the total number of garbage sequence models, for which the probability value is above a predefined value. Such a calculation of probability values for models is common use.  
10 Therefore the predefined probability value, which is used here to classify the garbage sequence model as a model representing the keyword or not, is determined empirically.

In accordance with a second aspect of the present invention, further  
15 - a path through the plurality of garbage models is detected, which matches best to a part of the spoken utterance,  
- a likelihood is calculated for that path, if the garbage sequence model is contained in that path  
- and wherein for assessing the part of the spoken utterance as  
20 the keyword to be recognized, that path through the plurality of garbage models is assumed as the garbage sequence model, when the likelihood is above a threshold.

For this, one garbage sequence model is required, which best represents the keyword. This garbage sequence model is  
25 determined and stored a-priori, before the recognition phase. If during the recognition phase, a path through the plurality of garbage models is detected, which matches best to a part of the spoken utterance then a following post-processing step is applied. In that post-processing step, a likelihood is  
30 determined, if the predefined garbage sequence model is contained in that path. If the likelihood is above a threshold, the path or a part thereof is assumed as the garbage sequence model. With that assumption the part of the spoken utterance is

assessed as the keyword to be recognized. Because only one garbage sequence model has to be stored, that recognition method according to the second aspect of the present invention causes less memory consumption and can therefore advantageously  
5 be applied, when the memory size is limited, like for example in mobile phones. Advantageously, because the threshold can be adjusted at any time for the needs, the recognition method according to that second aspect has a high flexibility.

10 Preferably the likelihood is calculated, based on the determined garbage sequence model, the detected path through the plurality of garbage models, and a garbage model confusion matrix, and wherein the garbage model confusion matrix contains the probabilities  $P(i|j)$  that a garbage model  $i$  will be recognized supposed a garbage model  $j$  is given.

15 Advantageously, the at least one garbage sequence model is determined, when a keyword model is created for a new keyword to be recognized. By this, the speech recognition method according to the first and the second aspect of the present invention is flexible, because the garbage model sequences are  
20 determined as soon as a new keyword is created. This is an advantage for speaker dependent recognition methods, where the keyword models are created from one or more utterances from one speaker, which in general is the user. Then the method is applied as soon as a new keyword is created from the user.

25 A further aspect of the present invention relates to a computer program product, with program code means for performing the recognition method according to the present invention, when the product is executed in a computing unit.

30 Preferably the computer program product is stored on a computer-readable recording medium.



In the following the advantages of the present invention will be apparent upon reading the following detailed description of the preferred embodiments and upon the following drawings where:

- 5        Fig.1 shows a finite state syntax for keyword spotting according to the first aspect of the present invention, Fig.2 shows a finite state syntax for determining a garbage sequence model according to the present invention, Fig.3 shows a mapping of a path through a plurality of garbage models to a garbage sequence model according to  
10        the second aspect of the invention, Fig.4 shows a finite state syntax for prior art keyword spotting, Fig 5 shows a block diagram of an automatic speech  
15        recognition device in a mobile equipment.

Automatic speech recognition is used to recognize one or more keywords from a spoken utterance. Therefore, the applied recognition method is depicted as a finite state syntax. Fig.4 shows a prior art finite state syntax for recognizing one  
20        keyword. Such a finite state syntax compares any part of an incoming utterance with models representing a keyword to be recognized. In Fig.4, a keyword model, created for the keyword to be recognized is shown as one path. Further a plurality of garbage models  $g_i$ , where  $i$  is an integer, is shown. For  
25        example, some garbage models represent speech events, like e.g. filled pauses "em" or lip smacks. Further garbage models represent other non-speech events, like background noise. To predefine the garbage models  $g_i$  it is important to have knowledge about the set of keywords, the acoustic environment  
30        in which the speech recognition is used, and the speech events to be covered by the garbage models. Additionally a further path is included in the finite state syntax, which is named

SIL-Model and represents a typical period of silence. As soon as the recognition is active, each incoming utterance or any part of the incoming utterance is matched to the stored models in the finite state syntax. For it, in the finite state syntax, a path through any of the predefined keyword-, SIL- and garbage- models is determined, which matches best to the incoming utterance. Here, a path can include only one of the models, or a series of the models. The keyword is recognized if the keyword model itself is included in the path.

In accordance with the principle concept of the present invention, a garbage sequence model is created, which also represents the keyword. This garbage sequence model then is used to assess the incoming utterances or a part thereof as the keyword to be recognized, if the garbage sequence model matches best to the incoming utterance or to the part of the utterance. The garbage sequence model is defined in the present invention as a series of consecutive garbage models  $g_i$ . Such a garbage sequence model is preferably created, based on the finite state syntax as depicted in Fig.2. Here, the finite state syntax for determining the garbage sequence model includes only a SIL-model and a plurality of garbage models  $g_i$ . The SIL-model is optional. The garbage models  $g_i$  are the same as used in the finite state syntax during the normal recognition phase. For the determination of the garbage sequence model, the finite state syntax as depicted in Fig.2, is applied to a keyword utterance, which represents the keyword to be recognized. Then that path through the plurality of garbage models  $g_i$  is selected, which matches best to the keyword utterance. This determined path, which is a series of consecutive garbage models  $g_i$ , is then used during the speech recognition phase to assess any part of an utterance as the keyword to be recognized. The creation of garbage sequence models according to the present invention can be used for speaker dependent and

speaker independent speech recognition. For speaker dependent speech recognition the keyword utterance, which represents the wanted keyword is speech, which is collected from one speaker. That speaker is usually the user of the mobile equipment, where the speech recognition method is implemented. For speaker independent speech recognition the keyword utterance is speech, which is collected from a sample of speakers. Alternatively, the keyword utterance is an already trained and stored reference model.

The method in accordance with the first aspect of the present invention is now described by an example, as depicted in Fig.1. Here the finite state syntax has one keyword model, one SIL-model, and a plurality of garbage models  $g_i$ . Further, exactly one garbage sequence model is used, which is created according to the present invention. In the present example the garbage sequence model consists of the series  $g_7$ - $g_3$ - $g_0$ - $g_2$ - $g_1$ - $g_5$  of consecutive garbage models, which are determined, based on the syntax as shown in Fig. 2. The finite state syntax, as shown in Fig.1, is then applied to an incoming utterance. With it, the hit rate is increased, because a keyword is recognized, if the part of the spoken utterance either matches best to the keyword model or to the determined garbage sequence model. Even if the method according to the first aspect of the present invention is described based on the finite state syntax as depicted in Fig.1, where exactly one garbage sequence model is used, the present invention is not limited to that example. Of course, a further number  $N$  of garbage sequence models can exist for each keyword to be recognized. With these further  $N$  garbage sequence models in addition to the first determined garbage sequence model, the hit rate is further increased. The total number  $N$  is limited, based on the probability that each of the  $N+1$  garbage sequence models represents the keyword. Therefore, for each of the determined garbage sequence models, a probability value is calculated. Then, those garbage sequence

models are selected as the total number  $N+1$  of garbage sequence models, for which the probability value is above a certain threshold. A typical threshold is assumed as a probability value, which is 90% from the maximal available probability value, wherein the maximal available probability value is the probability value for the best garbage sequence model. To limit the total number  $N+1$  of garbage sequence models to an operable amount, the total number  $N+1$  of used garbage sequence models should be limited to maximal 10.

Advantageously the determined garbage sequence models are privileged against any path through the plurality of garbage models. Particularly the series of consecutive garbage models, which determined the garbage sequence model, is always weighted higher than the same series of consecutive garbage models from the plurality of garbage models. Then the hit rate is increased, because as soon as a series of consecutive garbage models match best to the part of a spoken utterance, the garbage sequence model is selected and the part of the utterance is assessed as the keyword to be recognized. Even if the present invention is explained based on the finite state syntax for one keyword, the invention is also usable for more than one keyword. To privilege the garbage sequence model a penalty is defined for the garbage models from the plurality of garbage models. This then leads to a higher probability for the garbage sequence model, compared to an identical series through the plurality of garbage models.

A mapping from a path through a plurality of garbage models to the predefined garbage sequence model is depicted in Fig.3. Here, on the abscissa the determined garbage sequence model  $g_7-g_3-g_0-g_2-g_1-g_5$ , which matches best to the keyword model, is shown. A detected path through the plurality of garbage models, which matches best to the part of the incoming spoken utterance, is depicted on the  $t$  axis. The determined

garbage sequence model is already predefined, which for example is done according to the finite state syntax as shown in Fig.2. But contrary to the method in accordance with the first aspect, that garbage sequence model is not used directly to assess a part of an utterance as the keyword to be recognized. Rather, for recognition purposes, a prior art finite state syntax like that one shown in Fig.4 is used. In a first step, a path through the plurality of garbage models is detected, which best matches to the spoken utterance. Then, in a post-processing step, that detected path is compared with the predefined garbage sequence model. Therefore, a likelihood is calculated, that the predefined garbage sequence model is contained in the detected path. And finally, that path is assumed as the garbage sequence model, when the likelihood is above a certain threshold. When the path is assumed as the garbage sequence model, then the part of the spoken utterance is assessed as the keyword to be recognized. Also, the method in accordance with the second aspect of the present invention increases the hit rate. Contrary to the method in accordance with the first aspect, this method is more flexible, but it needs more computation effort. Here, for each keyword model, only one garbage sequence model has to be stored and the recognition process is post-processing computation. Based on Fig.3, the post-processing computation, where a keyword is assessed is now described in more detail. A soft comparison is applied by computing the likelihood, that the garbage sequence model is contained in the detected path through the plurality of garbage models. This likelihood is calculated for example by using a dynamic programming [Dynamic Programming; Bellman, R.E.; Princeton University Press; 1972] and a garbage model confusion matrix. At each point of the grid, which is shown in Fig.3, a probability is calculated, which describes the likelihood that the determined path matches with the predetermined garbage sequence model. Therefore the probabilities  $P(g_i|g_j)$ , where  $i \neq j$

and  $i, j$  are integer, which are known from the garbage confusion matrix are used as emission probabilities. Alternatively statistical models of higher order may be used as well. The transition probabilities for going from garbage model  $g_i$  at the  
5 time  $t$  to the garbage model  $g_j$  at the discrete time  $t+1$  are constant for all  $i, j, t$  and do not have to be considered in the search therefore. Also it is allowed either to remain in the same garbage model of the garbage sequence model from  $t$  to  $t+1$ , or to move to the next garbage model, or to skip a garbage  
10 model. Thus the dynamic programming search delivers the best probability, for the garbage sequence in the time interval from  $t_0$  to  $(t_0+M)$ , if the garbage sequence model was not exactly found in the path, as shown in Fig.3. In the post-processing step all possible paths through the grid network are calculated  
15 and the path with the highest probability is then used for the assessing step. In a final step the part of the spoken utterance is assessed as the keyword to be recognized, if the dynamic programming delivers a probability higher than a predefined threshold. Again also the method according to the  
20 second aspect of the present invention is not limited to the recognition of only one keyword. For more than one keyword the method is applied to each of the plurality of keywords.

The method in accordance with the principle concept of the present invention increases the hit rate. The hit rate is  
25 further increased with the both described aspects of the present invention. The method in accordance with the first aspect of the present invention is easy to implement and needs less computation effort. The method in accordance with the second aspect of the present invention is more flexible. The  
30 hit rate can also be increased when applying a method, which combines the features of the first and the second aspect of the present invention. Then, a part of the spoken utterance is assessed as the keyword, when in accordance with the first

aspect, the path directly matches best to one or more predefined garbage sequence models, or when in accordance with the second aspect, the path is assumed as the garbage sequence model. With it, the speech recognition method of the present invention is flexible and adaptable to the mobile equipment limitations, like e.g. limited memory size in that mobile equipment, where the method is implemented.

Fig 5 shows a block diagram of an automatic speech recognition device 100 in a mobile equipment, like e.g. a mobile phone. The central parts of the speech recognition device 100, which are arranged as several parts (as shown) or as one central part, are: a pattern matcher 120, a memory part 130 and a controller part 140. The pattern matcher 120 is connected with the memory part 130, where the keyword models, the garbage models, the SIL-model and the garbage sequence models can be stored. The keyword models, the SIL-models and the garbage models are created according to well known prior art techniques. The garbage sequence models are determined in accordance with the present invention, as described above. The controller part 140 is connected to the pattern matcher 120 and to the memory part 130. The controller part 140, the pattern matcher 120 and the memory part 130 are the central parts, which carry out any of the methods for automatic speech recognition of the present invention. An utterance, which is spoken from a user of the mobile equipment, is transformed from a microphone 210 in an analog signal. This analog signal is then transformed from an A/D converter 220 in a digital signal. That digital signal is then transformed from a pre-processor part 110 in parametric description. The pre-processor part 110 is connected to the controller part 140 and the pattern matcher 120. Based on a finite state syntax according to the present invention, the pattern matcher 120 compares the parametric description of the spoken utterance with the models, which are stored in the memory part 130. If the parametric description

from at least a part of the spoken utterance matches to one of the stored models in the memory part 130, an indication of what is assessed as to be recognized is given to the user. That indicated recognition result is conveyed to the user by a  
5 loudspeaker 300 or on a display (not shown) of the mobile equipment.

Contrary to speech recognition devices, known from prior art, the automatic speech recognition device according to the present invention, also assesses any part of the spoken  
10 utterance as a keyword to be recognized, if that part matches best to at least one of the determined and in the memory part stored garbage sequence models. With that, the hit rate is increased.

15



## CLAIMS

1. Method for recognizing a keyword from a spoken utterance,  
with at least one keyword model and a plurality of garbage  
models, **wherein**

5 a part of the spoken utterance is assessed as the keyword to  
be recognized, if that part matches best either to the  
keyword model or to a garbage sequence model,  
and wherein the garbage sequence model is a series of  
consecutive garbage models from that plurality of garbage  
10 models.

2. The method according to claim 1, **wherein** the garbage sequence  
model is determined

- by comparing a keyword utterance, which represents the  
keyword to be recognized, with the plurality of garbage  
15 models and

- detecting the series of consecutive garbage models from  
that plurality of garbage models, which match best to the  
keyword to be recognized.

3. The method according to claim 1 or 2, **wherein**

20 - the determined garbage sequence model is privileged against  
any path through the plurality of garbage models.

4. The method according to any of the claims 1-3, further

- determining a number (N) of further garbage sequence  
models, which also represent that keyword to be recognized,  
25 and

- assessing the part of the spoken utterance as the keyword  
to be recognized, if that part of the spoken utterance  
matches best to any of that number (N) of garbage sequence  
models.

30 5. The method according to claim 4, **wherein** the total number  
(N+1) of garbage sequence models are determined:

- by calculating for each garbage sequence model a  
probability value and

- selecting those garbage sequence models as the total number (N+1) of garbage sequence models, for which the probability value is above a predefined value.

6. The method according to any of the claims 1-5, further

- 5     - detecting a path through the plurality of garbage models, which matches best to the spoken utterance,  
      - calculating a likelihood for that path, if the garbage sequence model is contained in that path and  
      - wherein for assessing a part of the spoken utterance as the  
10    keyword to be recognized, that path through the plurality of garbage models is assumed as the garbage sequence model, when the likelihood is above a threshold.

7. The method according to claims 6, **wherein**

- the likelihood is calculated based on the determined  
15    garbage sequence model and the detected path through the plurality of garbage models and a garbage model confusion matrix, and  
      - wherein the garbage model confusion matrix contains the probabilities  $P(i|j)$  that a garbage model  $i$  will be  
20    recognized supposed a garbage model  $j$  is given.

8. The method according to claim 7, **wherein** the likelihood is calculated with dynamic programming techniques.

9. The method according to any of the claims 1-8, **wherein** the at least one garbage sequence model is determined, when a  
25    keyword model is created for a new keyword to be recognized.

10. The method according to any of the claims 1-9, **wherein** the keyword utterance is speech, which is collected from one speaker.

11. The method according to any of the claims 1-9, **wherein** the  
30    keyword utterance is speech, which is collected from a sample of speakers.

12. The method according to any of the claims 1-9, **wherein** the keyword utterance is a reference model.

13. A computer program product with program code means for performing the steps according to one of the claims 1 to 12  
5 when the product is executed in a computing unit.

14. The computer program product with program code means according to claim 13 stored on a computer-readable recording medium.

15. An automatic speech recognition device 100, implemented  
10 the method according to any of the claims 1-12, including  
- a pre-processing part (110), where a digital signal from an utterance, spoken into a microphone (210) and transformed in an A/D converter 220 is transformable in a parametric description;

15 - a memory part (130), where keyword models, SIL models, garbage models and garbage sequence models are storable;  
- a pattern matcher (120), where the parametric description of the spoken utterance is comparable with the stored keyword models, SIL models, garbage models and garbage sequence  
20 models;

- a controller part (140), where in combination with the pattern matcher (120) and the memory part (130), the method for automatic speech recognition is executable.

16. A mobile equipment, with an automatic speech recognition  
25 device according to claim 15, wherein the mobile equipment is a mobile phone.

1/3

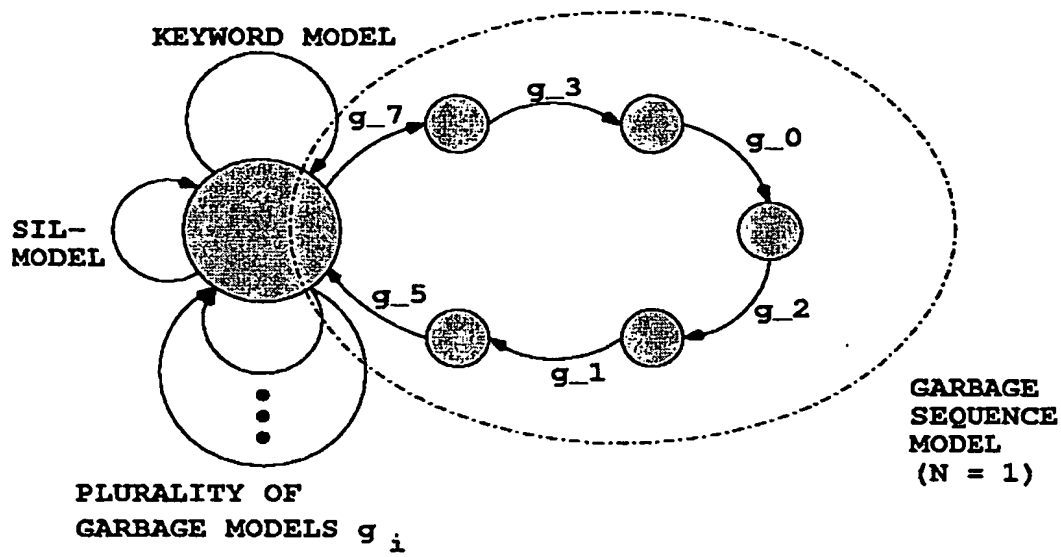


FIG 1

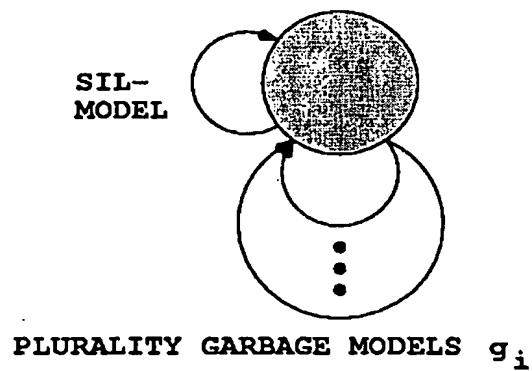


FIG 2

SUBSTITUTE SHEET (RULE 26)

2/3

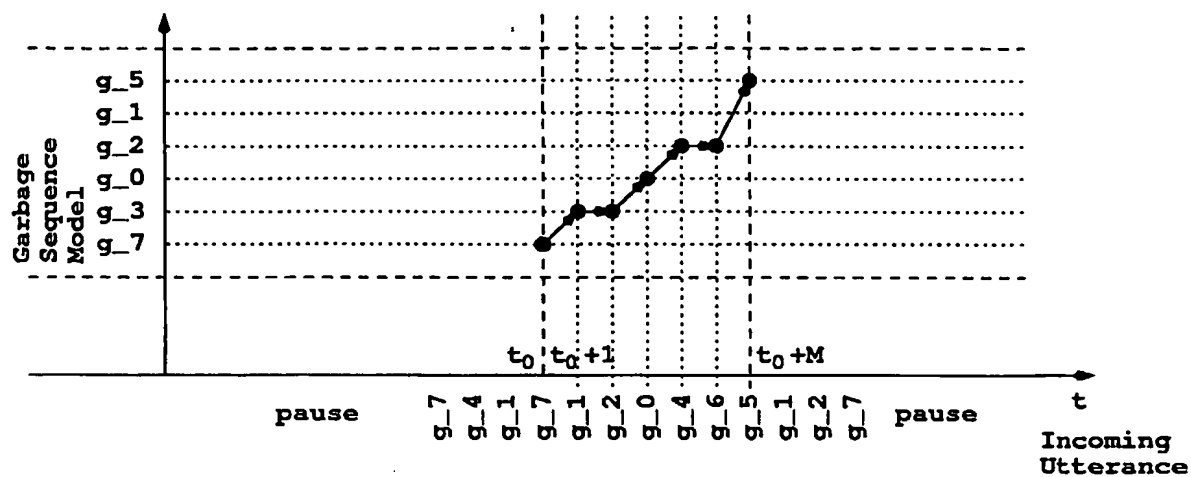
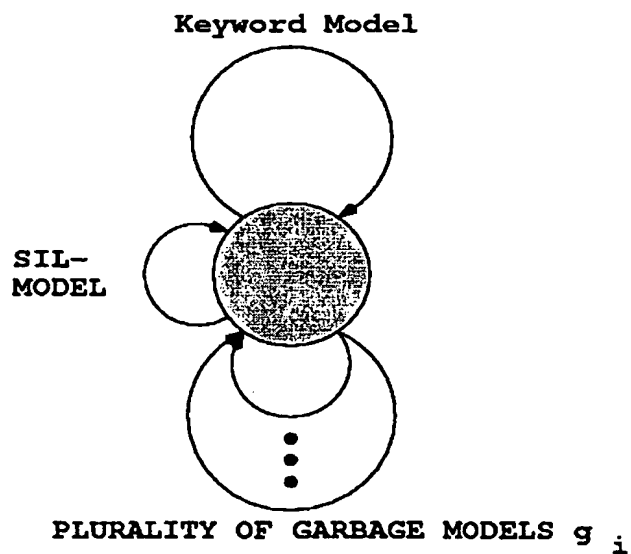


FIG 3



Prior Art

FIG 4

SUBSTITUTE SHEET (RULE 26)

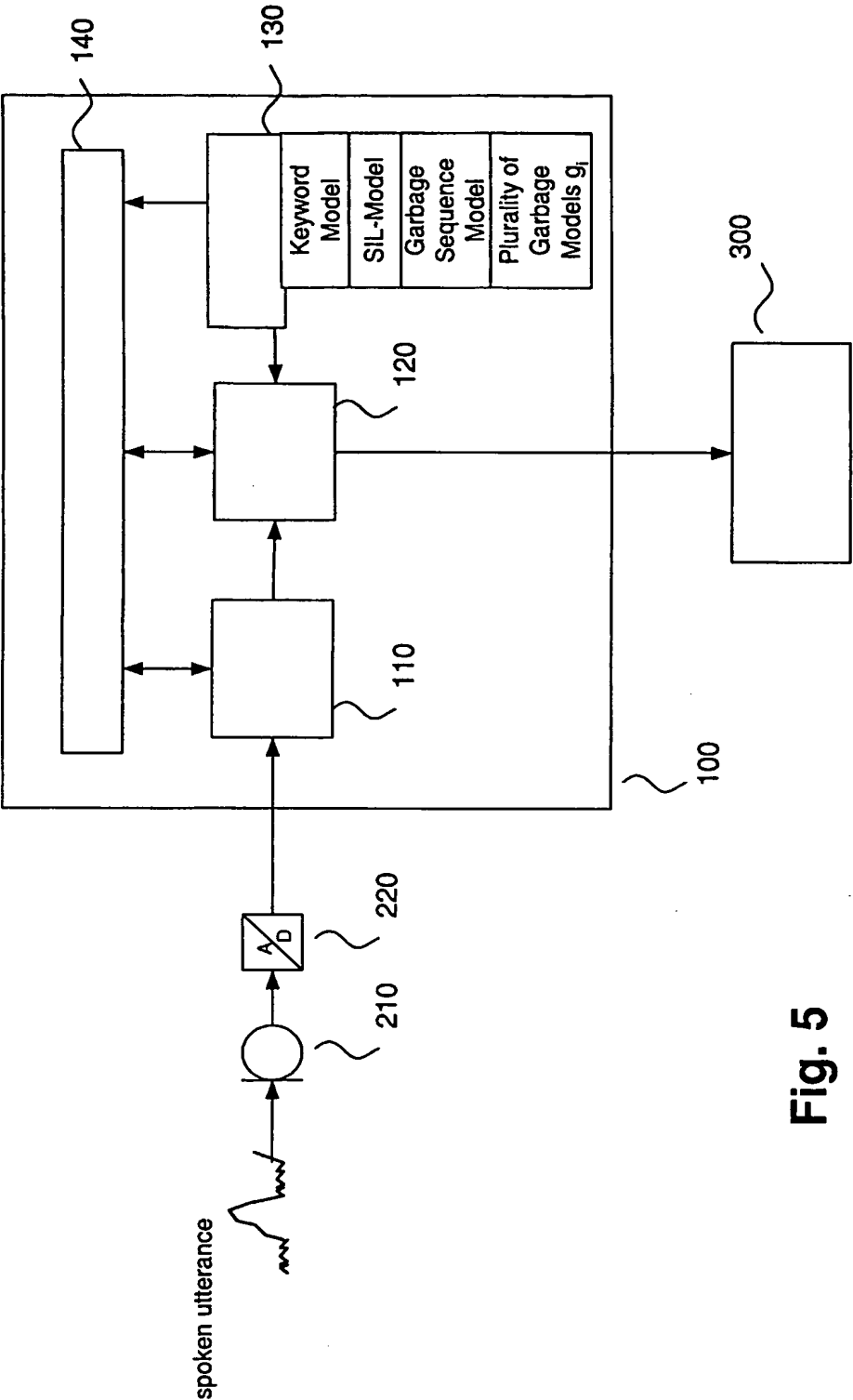


Fig. 5

# INTERNATIONAL SEARCH REPORT

International Application No.

PCT/EP 02/08585

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G10L15/20 G10L15/18

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YAMADA M ET AL: "ACTIVE/NON-ACTIVE WORD CONTROL USING GARBAGE MODEL - UNKNOWN WORD RE-EVALUATION IN SPEECH CONVERSATION" ICSLP 94: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP), vol. 2, 18 - 22 September 1994, pages 823-826, XP000855376 YOKOHAMA: ASJ, JP abstract * section "2.2 Unknown word detection using garbage model" *	1, 2, 4, 9-11, 13-16
A	EP 1 215 654 A (SONY INTERNAT EUROP GMBH) 19 June 2002 (2002-06-19) abstract; figure 1 --- -/-	1, 10, 11, 13-16

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*G\* document member of the same patent family

Date of the actual completion of the international search

2 April 2003

Date of mailing of the international search report

16/04/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Quélavoine, R

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 02/08585

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>EP 1 215 660 A (ERICSSON TELEFON AB L M)  19 June 2002 (2002-06-19)  abstract</p> <p>-----</p>	1, 13-16



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP 02/08585

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 1215654	A	19-06-2002	EP 1215654 A1	19-06-2002
			JP 2002229587 A	16-08-2002
			US 2002116193 A1	22-08-2002
<hr/>				
EP 1215660	A	19-06-2002	EP 1215660 A1	19-06-2002
			AU 3323802 A	24-06-2002
			WO 0249006 A1	20-06-2002
			US 2002091526 A1	11-07-2002
<hr/>				